AD_____

Award Number: W81XWH-04-1-0672

TITLE: Prediction of Breast Cancer Risk by Whole Genome Profiling

PRINCIPAL INVESTIGATOR: Jeff Boyd, Ph.D.

CONTRACTING ORGANIZATION: Sloan-Kettering Institute for
Cancer Research
New York, NY 10021

REPORT DATE: September 2005

TYPE OF REPORT: Final

PREPARED FOR: U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

REPORT DOCUMENTATION PAGE

Form Approved
OMB No. 0704-0188

1. REPORT DATE	2. REPORT TYPE	3. DATES COVERED
01-09-2005	Final	6 Aug 2004 – 5 Aug 2005

4. TITLE AND SUBTITLE

Prediction of Breast Cancer Risk by Whole Genome Profiling

5a. CONTRACT NUMBER

5b. GRANT NUMBER
W81XWH-04-1-0672

5c. PROGRAM ELEMENT NUMBER

6. AUTHOR(S)

Jeff Boyd, Ph.D.

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES)

Sloan-Kettering Institute for
Cancer Research
New York, NY 10021

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING / MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Medical Research and Materiel Command
Fort Detrick, Maryland 21702-5012

10. SPONSOR/MONITOR'S ACRONYM(S)

11. SPONSOR/MONITOR'S REPORT NUMBER(S)

12. DISTRIBUTION / AVAILABILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES

14. ABSTRACT

The purpose of this Concept Award project was to determine the feasibility of genotyping a population of Ashkenazi Jewish (AJ) breast cancer cases and controls using high density single nucleotide polymorphism (SNP) microarrays, with the eventual goal of testing the hypothesis that most or all of breast cancer cases represent a complex polygenic disease, susceptibility to which can be assessed using high density SNP genetic profiling. Work under this award was accomplished in two phases. In the first, 100 AJ controls were genotyped using Affymetrix 500K and 100K SNP arrays; these experiments proved that the work was feasible and that high quality data could be generated using these reagents, and provided the data to construct the first detailed haplotype map of the AJ population. Analysis and preparation of these data for presentation are now underway. In the second phase, an additional 300 AJ cases and 200 controls will be genotyped using Affymetrix 500K and 100K SNP arrays, for a total of 300 cases and 300 controls. The DNA specimens have been collected for phase two, and genotyping of these samples is underway.

15. SUBJECT TERMS
Breast cancer, genetic susceptibility, hereditary, genotyping, single nucleotide polymorphisms, microarray

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	18. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			USAMRMC
U	U	U	UU	8	19b. TELEPHONE NUMBER *(include area code)*

Table of Contents

Cover.....1

SF 298.....2

Table of Contents..... 3

Introduction.....4

Body.....5

Key Research Accomplishments.....6

Reportable Outcomes.....7

Conclusions.....8

References.....N/A

Appendices.....N/A

Introduction

It is reasonably well established that about 7% of all breast cancer cases are associated with autosomal dominant genetic predisposition, a substantial fraction of which are associated with inherited mutations in *BRCA1* or *BRCA2*. The remaining 90+% of all breast cancers have heretofore been presumed to occur "sporadically", that is, as a result of acquired mutations in genes that occur in breast cells as a result of lifestyle, hormonal, or environmental exposures, or spontaneously as a result of faulty DNA replication and repair. While all of these factors may indeed contribute somewhat to the development of "sporadic" breast cancer, evidence is accumulating to support a hypothetical model in which genetically susceptible women contribute a high proportion, and perhaps the majority, of overall breast cancer incidence (1). This evidence includes the observed high constant incidence of breast cancer in the contralateral breast, in twins, and in other relatives of women with breast cancer (2), and a population-based segregation analysis demonstrating a log-normal distribution of risk in the population (3), which together suggest that a major proportion of breast cancers occur in a susceptible minority of the population. Unlike those cases associated with dominant, highly penetrant genes such as *BRCA*, however, this model holds that most breast cancers may be classified as complex genetic diseases, resulting from a combination of constitutional genetic variants affecting a large number of different genes. The purpose of this Concept Award project was to determine the feasibility of genotyping a population of Ashkenazi Jewish (AJ) breast cancer cases and controls using high density single nucleotide polymorphism (SNP) microarrays, with the eventual goal of testing the hypothesis that most or all of breast cancer cases represent a complex, polygenic disease, susceptibility to which can be assessed using high density SNP profiling.

Body

The specific aims of this pilot study were to: 1) Genotype 300 AJ breast cancer cases and 300 AJ controls using high density (500K + 100K) SNP microarrays; 2) Perform class prediction using the SNP data and a variation of logic regression (see below) to determine whether cases may be distinguished from controls; and 3) Assess the likely statistical significance of the genetic predictor using multiple validation techniques.

The human study material consisted of peripheral blood lymphocyte DNA from breast cancer cases and controls drawn from an existing DNA bank in the PI's laboratory, which consists of 1,250 incident cases of invasive breast cancer and 1,250 controls (post-menopausal women with no personal or family history of cancer); all of the cases and controls were obtained at a single institution with informed consent under an IRB-approved protocol. *These existing specimens were anonymized for this study, which is thus exempt under 32 CFR 219.101(b)(4).* To minimize biological and genetic heterogeneity, the subset of cases and controls used for this study were all postmenopausal and of Ashkenazi Jewish ethnicity. These DNA samples are to be analyzed using the Affymetrix GeneChip Human Mapping 100K and 500K SNP Arrays (obtained at this institution through an Early Access agreement with Affymetrix), which allows for the interrogation of approximately 600,000 well-characterized SNPs distributed throughout the genome. The genotyping is performed at a Genomics Core Facility at this institution that has all necessary equipment and software for this type of analysis. A variation of logic regression will be used for statistical analysis. Logic regression is a new adaptive regression methodology that attempts to construct predictors as Boolean combinations of binary covariates. This algorithm was recently modified to deal with SNP data (4). The predictors that are found may be interpreted as risk factors for the disease (breast cancer in this case). Statistical significance of these risk factors is assessed using techniques like cross-validation, permutation tests, and independent test sets. This technique has been used successfully to uncover the complex genetic basis of several diseases based on the analysis of SNP data.

Successful proof-of-concept would represent a substantial advance toward the ability to predict women at substantial and insignificant risks for breast cancer, which would be expected to have a major impact on breast cancer screening and prevention paradigms currently adhered to.

Key Research Accomplishments

Work under this award was accomplished in two phases. In the first, 100 AJ controls were genotyped using Affymetrix 500K and 100K SNP arrays; these experiments proved that the work was feasible and that high quality data could be generated using these reagents, and provided the data to construct the first detailed haplotype map of the AJ population. Analysis and preparation of these data for presentation are now underway.

In the second phase, an additional 300 AJ cases and 200 controls will be genotyped using Affymetrix 500K and 100K SNP arrays, for a total of 300 cases and 300 controls. The DNA specimens have been collected for phase two, and genotyping of these samples is underway.

Reportable Outcomes

None to date.

Conclusions

1. Genotyping of DNA obtained from human blood samples using Affymetrix 500K and 100K SNP Arrays is feasible and yields high quality data with very high call rates (>98%).

2. Genotyping of 100 AJ control subjects using the 500K and 100K SNP arrays has been completed, and the first comprehensive haplotype map of the AJ population is under construction using these data.

3. Completion of the project, which will involve the genotyping of a total of 300 AJ breast cancer cases and 300 controls, is underway. Statistical analysis of these data within the calendar year 2005 is expected to yield proof of concept of the underlying hypothesis of this project, as outlined in the Introduction to this progress report.